DOCUMENT RESUME

ED 141 467 UD 017 101

AUTHOR Spencer, Thelma L.
TITLE Ethnic Minorities and National Standardized Testing.
PUB DATE 75
NOTE 21p.; Paper presented at the Third Annual Conference of the Society of Ethnic and Special Studies (Atlanta, Georgia, November 12-15, 1975)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Blacks; Educational Testing; *Ethnic Groups; Group Intelligence Testing; *Historical Reviews; Intelligence Tests; *Minority Groups; National Norms; *Standardized Tests; *Testing Problems; Test Interpretation; Test Results; Test Validity

ABSTRACT

This paper provides a historical review of national standardized testing and its relation to ethnic and racial minorities. In the pre-World War I period, psychological testing was conducted on the large masses of immigrants that were arriving in the U.S., and on black and white army draftees. Generally these tests showed that black draftees and refugees scored lower than the white middle class. No one in this period questioned the appropriateness of these tests for people whose backgrounds, language, and life styles were different than the majority of the population. But then, as now, the cultural bias inherent in such tests served the very useful purpose of labeling. Until recently race was still not considered to be a factor in test sampling. Furthermore it is only recently that there are provisions made in test administration for children whose first language is not English. Many who presently oppose the use of standardized tests with minority and ethnic groups and blacks do so out of the belief that such tests, especially intelligence tests, cannot be divorced from the cultural frame within which the individual exists, lives, and learns. While it may be argued that procedures of test development are less biased now than forty, twenty, or ten years ago, two problems remain to be solved: the development of test content representative of a pluralistic American society, and more appropriate use of tests and test results. (Author/AM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED141467

ETHNIC MINORITIES AND NATIONAL STANDARDIZED TESTING

by

Thelma L. Spencer
Educational Testing Service
Princeton, New Jersey

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

UD017101

Third Annual Conference of the
Society of Ethnic and Special Studies
Atlanta, Georgia
November 12-15, 1975

©1977 Thelma L. Spencer

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Thelma L. Spencer

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

Ethnic Minorities and National Standardized Testing

From the day in the distant past when the forebears of today's Native Americans crossed the land bridge from Asia into North America until 1875 there was no law prohibiting immigration into this country. In that year a law was passed that, even with subsequent amendments, established no numerical limitation on immigration, but did prohibit the entry of coolies, convicts, and prostitutes, among other excluded classes of undesirables.

According to one observer, control over immigration developed slowly and at first by the gradual addition of new excluded classes. While there were regulations about areas from which immigration was precluded to assure an appropriate racial balance (China, e.g.), there were no discriminations drawn among the various countries of northern and western Europe that provided the bulk of immigrants. (Kamin, p. 15) The truth of the matter is that the needs of the nation's rapidly expanding economy could be filled with immigrants. It was obvious to any astute observer that the place of respective ethnic groups in the industrial class structure was characterized by the ethnic identity of the group. This identity, formed by the privileges and privations of the group in the nation of its origin, determined one's place in the social class structure of the United States.

Kamin's comments in this regard are descriptive of how this operated. He writes

> With the turn of the century, the "New Immigration" from southeastern Europe began to assume massive proportions. The English, Scandanavian, and German stock which had earlier predominated was now outnumbered by a wave of Italian, Polish, Russian, and Jewish immigrants. The popular press and the literary magazines of the period were filled with articles questioning the assimilability of the new and exotic ethnic breeds. (Kamin, p. 16)

The fear engendered by such articles was reflected in insistent demands for some measure of control over the quality of immigrants, and the response led to the development and implementation of certain practices whose outcomes continue to influence social and educational institutions and policy.

This "quality control" first took the form of a demand for a literacy test, and eventually the use of the Binet test and supplementary performance tests to assess the "innate intelligence" of the great mass of average immigrants. The results of Binet tests given to immigrants on Ellis Island in 1912 established conclusively for those administering the tests that 83% of the Jews, 80% of the Hungarians, 79% of the Italians, and 87% of the Russians seeking entry into the U.S. were "feebleminded." (Kamin, ibid.)

Then, as now, the cultural bias inherent in such tests served the very useful purpose of what Jane Mercer so aptly calls "labeling." No one in the pre-World War I period questioned the appropriateness of those tests for people whose backgrounds, languages, and life styles

were non-American. The congenial climate in which the psychological testing movement was nurtured was further enhanced by two events: the importation of the Binet Test from France and the United States' entry into the war.

It was during World War I that a group of psychologists under Robert Yerkes developed a group intelligence test to provide mental assessments of army draftees. Knowledge of test results was supposed to aid in classifying the draftees for army jobs. Yerkes and his colleagues came up with two measures of intelligence--a written group intelligence test, the Alpha; and a supplementary test for illiterates, the Beta, that was a non-verbal or performance battery. In addition, pantomime instructions were given to non-English speaking testees taking the Beta form.

Approximately two million men took the tests but, even though their use was widely publicized, test results were not much used for job placement of draftees. However, this vast store of data made it possible for later researchers to analyze the elements entering into something one of them called "American intelligence." (Brigham, p. xxi) In his view, the intelligence test scores of the native born, the foreign born, and the Negro afforded an opportunity for a "national inventory of our own mental capacity, and the mental capacity of those we have invited to live with us." (Brigham, p. xx)

Carl C. Brigham's analysis of these data in a 1923 report entitled, A Study of American Intelligence, revealed to him what he chose to identify as a decline in the intelligence of immigrants arriving after 1902. He attributed this decline to two factors: the change in the races migrating to this country, and the sending of lower and lower representatives of each race. According to his interpretation, the decline of immigrant intelligence paralleled precisely the decrease in the amount of "Nordic blood," and the increase in the amount of "Alpine" and "Mediterranean" blood in the immigrant stream.

As for the Negroes in the sample, he concluded that his results showing the marked intellectual inferiority of this group were corroborated by previous investigations by researchers who had used psychological tests on white and Negro groups. When a difference was found between northern and southern Negroes in the sample, he ascribed the northern Negro's superior intelligence measurements to three factors: first, the greater amount of educational opportunity, which affected scores on existing intelligence tests; second, the greater amount of admixture of white blood; third, the operation of economic and social forces such as higher wages, better living conditions, identical school privileges, and a less complete social ostracism, tending to draw the more intelligent Negro to the North.

Brigham's analysis and interpretation of the data about the intelligence of people from almost every immigrant European racial group

and the Negroes whose forefathers arrived on these shores in 1619 provided the basis for the passage of the Immigration and Naturalization Act of 1924. The Act, setting quotas on the numbers of immigrants on the basis of the proportionate representation of their particular ethnic group in the U.S. Census of 1890, was in effect a retraction of the invitation to "live with us" as far as certain racial and ethnic groups were concerned.

I have chosen to cite the Brigham study and the political use of his analysis and interpretation to establish national public policy as one of the earliest misuses of data derived from administering a psychological test. It is paradoxical that a man as bright as Brigham was reputed to be could not (or would not) see the fallacy in his reasoning nor the chasm between presumed cause and effect. For example, his analysis revealed what he described as the relationship between length of residence in the United States and increase in the average test score. It was to him remarkable that there was such a graphic relationship between these two factors. He wrote

> If our results reflect another factor independent of intelligence, which might be designated 'the better adaptation of the more thoroughly Americanized group to the situation of the examination,' we have no means of controlling this factor. Ultimately, the validity of our conclusions from this study rests on the validity of alpha, beta, and the individual examinations. It is sometimes stated that the examining methods stressed too much the hurry-up attitude frequently called typically American. The adjustment to test conditions is a part of the intelligence test. We have, of course, no other measure of adjustment aside from the total score on the examinations given. If the tests used included some mysterious type of situation

> that was 'typically American,' we are indeed fortunate, for this is America, and the purpose of our inquiry is that of obtaining a measure of the character of our immigration. Inability to respond to a 'typically American' situation is obviously an undesirable trait. (Brigham, pp. 93, 96)

Brigham used his interpretation to demonstrate the correlation between test score and years of residence in America. By pooling across all countries of origin, he showed that immigrants who had been in the United States 16 to 20 years before being tested were as bright as native-born Americans; and that immigrants who had been in America only zero to five years when tested were virtually feebleminded. His racist views actually led him to make a quantum leap into assuming that the tests actually measured native or inborn intelligence. Oddly enough, though, this distorted reasoning process did not lead him to similar conclusions about Negroes whose ancestors had been in America since the early 17th century. It would follow, if his conclusion had validity, that the Negroes in America would be at or near the top of his intelligence totem pole if years of longevity in residence were the criterion.

The dissemination of the results of Brigham's study and the work of Terman, Thorndike, McDougall, Hirsch, and others focused attention on the possibility of using group intelligence tests to measure learning ability and to form instructional groups on the basis of the results. Scores on group intelligence tests and, a few years later,

on standardized achievement tests became the measures on which were based most of the grouping practices initiated in the nation's schools. The pernicious effects of such grouping were visited upon Negroes and those who lived in homes where Italian, Hebrew, Greek, Polish, Slavic, or Spanish, rather than English, was the spoken language.

The interests of all, including those for whom he believed himself to be spokesman, would have been served more adequately had he been less biased in his interpretation of the results of the testing program because he came later on to have a great deal of influence in the area of standardized test development and the field of educational testing. An illustration of this would be the criticism of the use of standardized tests with children and youth in the nation's schools today, critics contending that the tests do not reflect the pluralistic nature of our society. Many who presently oppose the use of standardized tests with minority and ethnic groups and with blacks do so out of the conviction that such tests, especially intelligence tests, cannot be divorced from the cultural frame within which the individual exists, lives, and learns. In addition, and equally important, the tests are considered inappropriate for children who come from bilingual or non-standard English backgrounds. What is true in 1975 was just as true a half century ago.

There were no provisions made in test administrations for children whose first (and often preferred) language was something other than English. No directions were given in pantomime, and no performance tests were substituted to determine the intellectual abilities of children from divergent ethnic or racial backgrounds.

Tests of every kind were developed, standardized, and widely used with little thought given to their relevance and appropriateness for certain groups of people, or to the consequences of decisions made on the basis of test performance. In the normal and accepted routine of test standardization it is entirely possible that population samples did not even include children from first and second generation ethnic group families, black children from inner city urban schools, or any children from rural backgrounds. It would be helpful, therefore, for us to take a look at the sampling procedures used in standardization of a widely used elementary level achievement test. (Lennon, p. 22ff.)

A statistical model of the kinds of school systems desired is developed. This model includes the location of school systems, the national distribution with respect to socioeconomic status as defined by median monthly income, cultural level as reflected by average schooling of the adult population of the communities, and size of school system. School systems in this stratified sample are invited to participate. If a school system declines the invitation, it is replaced in the

sample with another matched on these characteristics. With the exception of very large school systems where a sample of students is tested, every child in participating systems is included in the standardization administration.

Until recently race was not considered to be a factor in selecting either entire school districts or selected samples within school districts. As one test publisher expressed it, test developers were not particularly sensitive to race as a dimension of the sampling matrix. The consequences of this attitude can be stated numerically to furnish an idea of how this attitude impacts on one group--blacks--especially. It is important to keep in mind that it is believed that stratified sampling gives one a group that is supposedly representative of the total population being sampled.

In a 1963 re-standardization of the Metropolitan and the Stanford Achievement Tests, and the Metropolitan Readiness tests, there were 18,000 Negro children in the total sample of 350,000 taking the Metropolitan Achievement tests, or about five per cent. The publisher claimed that his company's operation was color blind, and that the standardization group actually included anywhere from six to seven to nine per cent non-white. If this had been the case, he would have had to include 3,500, 7,000, or 14,000 more non-white children just to reach the representative sample he claimed to have.

Though important, this is not as important as another fact for our purposes. What is important is that test publishers rely on Census data preponderantly to determine the socioeconomic position of the children being tested, and the average level of schooling of the adult population, or the cultural index. It is apparently assumed that use of Census data helps test publishers establish more accurate samples for their statistical models. That this is not necessarily true is evidenced by the testimony given by the publisher of the Metropolitan in the case of Hobson vs. Hansen. (Lennon, p. 118.)

In the case of the Metropolitan standardization, the Census data showed that there were 35,465,000 children in the United States between the ages of five and 14, the years which include elementary school children with whom the Metropolitan test is used. The 4,364,000 Negro children comprised 12.3 per cent of the total. It does not appear that the standardization group was as representative as the publisher believed it to be. The percentage of Negro children actually included in the sample looks even worse if Mercer's findings are any indication of what happens to many non-white children in the schools of this nation. She found that in California many non-white children were labeled as slow learners and placed in EMR classes on the basis of performance on mental ability tests given to predict their ability to succeed in a regular school program. Spanish-surnamed children and black children were represented in EMR

classes in much greater proportion than their incidence in the total school-age population. (Mercer, p. 261.) If these findings can be extrapolated to other school systems in the United States many of the 4,364,000 Negro children were not present in regular classrooms to even be included in the 1963 Metropolitan re-standardization.

The publishers of the Metropolitan and Stanford Achievement Tests were not alone in their use of questionable procedures for selecting groups for standardization administrations. In the spring of 1960 the Houghton Mifflin Publishing Company released the L-M Revision of the Stanford-Binet Intelligence Test. The 1960 test, combining the so-called best features of the 1937 forms L and M, and presenting an improved format and packaging, had no representation of minority groups in the sample population. This was in accordance with the philosophy of the test developers that the inclusion of small samples of minority groups would add much to the variability of the test and very little to its representativeness. As a result, no Negro children were included in the normative sample even as the 1960 Census revealed that there were 1,110,393 Negro elementary school children living in the states of Alabama, Florida, Georgia, South Carolina, and Tennessee. The Negro child, especially the child from the southeastern part of the country, was at a disadvantage in taking the test because it included many white culturally-oriented items to which he had little or no exposure, and because it had sampled a population that included no Negroes.

By early 1971 there was reason to believe that the original norms derived in 1937 were no longer completely appropriate for the United States population of the 1970's which was becoming culturally and ethnically more diverse, especially in the urban areas. The norms for the Stanford-Binet were, therefore, updated in a testing program carried out during the 1971-72 school year. New norms tables were developed for the testing instrument which remained unchanged in content, administration, and scoring.

A representative sample of approximately 100 subjects was tested at each age level of the test, about 2,100 subjects in all. A technique for stratifying each age sample to ensure proportionate representation of all ability levels was developed, the stratifying variable being test scores obtained from a large population of students who participated in a recent national standardization of the Cognitive Abilities Test, a group mental ability test. This was the primary stratifying variable though consideration was also given to geographic location and to type and size of the communities selected for participation. All testing was done by professionally trained and certified Stanford-Binet testers, and in order to ensure their availability, preference was given to the selection of communities to be sampled to those located within a radius of 50 miles of a university or established testing center with such a professional complement.

The norming of the CAT had involved approximately 20,000 individuals in each age group tested. These samples were chosen from a large number of communities stratified in terms of size, geographic region and indices of the economic status of the community. These original groups were further stratified to provide appropriate norming samples for the Stanford-Binet. The S-B normative testing was carried out in seven locations: Placer County, California; Denver, Colorado; Shelton, Connecticut; Valparaiso, Indiana; Fort Worth and Houston, Texas; and Salt Lake City, Utah.

The test performance data revealed genuine, important and highly relevant shifts in performance characteristics, changes of a systematic nature from the pre-school years to the elementary school years and again to the secondary school years, necessitating changes in the norms tables. Caution was advised in interpreting the shift from the 1930's to the 1970's because the norms established nearly forty years before had been clearly and explicitly limited to the white population and to the presumably English-speaking population. Blacks, Mexican Americans and Puerto Rican Americans had not been included in the original sample. The 1970 CAT norming group was chosen by community without regard to race or language background, communities in which a proportion of Black and Spanish-surnamed persons would be found. Similar factions were found in the subsample of communities in which the Binet testing was carried out. Black and

Spanish surnamed children were tested in the 1972 Binet testing, but only if English was the primary language spoken in the home. It is the belief of the test developers that the resulting Binet norms are inclusive of the population of the United States without regard to racial or national origin. (S-B Test Manual, pp. 353-360.)

The procedures described in the literature about the 1972 standardization program of the Stanford-Binet failed to consider two points which I believe to be very important to one attempting to put all of this into some meaningful context. The first has to do with the communities in which the testing was done. It hinges upon the same assumption made by the developers of the Metropolitan test referred to above. The re-standardization of that testing program was conducted on the presumption that the sample was representative of the total population of school children to whom the test might be given, irrespective of race. We found that this was not true because while Negro children comprised over 12% of children in the United States between the ages of five and 14, they comprised only five per cent of the Metropolitan sample group.

With regards to the 1972 Stanford-Binet sampling base, it is not necessary that one have precise statistics to ascertain that this group is not as representative as claimed. The communities involved in the 1970 CAT norming leave something to be desired if ethnic representativeness is the criterion. For example, there are fewer than

7,000 Blacks in the State of Utah, thus it would be almost impossible for Salt Lake City to furnish an adequate percentage of that group's representation. On the other hand, with respect to Spanish-surnamed representation, one would not expect to find any significant percentage in such locales as Shelton, Connecticut (1970 population of 27,165) or Valparaiso, Indiana (1970 population of 20,020).

The fact that Denver, Houston and Fort Worth, with large numbers of Mexican American residents, were included in the CAT and S-B norming does not mean that children from this ethnic group were adequately represented in the norming samples, especially at the upper age levels, because they begin to drop out of school in large numbers at about age 13. One study revealed that between grades 7 and 12, about 65% of Mexican American children in Texas drop out of school. (Carter, p. 27) There were no comparable data readily available on the drop out rate of black students in that geographical region so one may only speculate about the extent to which this group of students was represented in the upper age levels of the S-B norms table. The second point that one must consider, then, is the absolute numbers of Black or Mexican American school age children enrolled and in attendance in any of the selected communities, and the degree to which their incidence in the school population is representative of the incidence of their ethnic group in the total population.

The use of sampling techniques as briefly described here calls into question the appropriateness of routine use of standardized test interpretive norms and material with non-white children. In addition, one must question the use of these instruments with children of any color who come from homes located in the rural and Appalachian regions of American which meet the criteria of poverty. Such children are not to be found in the communities in which the 1970 CAT and the 1972 S-B were administered. One may conclude, therefore, that to use these tests to make educational placement decisions about children from ethnic, minority or socioeconomic groups whose backgrounds diverge from that identified as white middle-class is to deny these children a fair assessment and evaluation of their abilities, achievements, and potentials.

While it may be argued that procedures of test development are less biased now than they were forty or twenty or even ten years ago, two very basic problems remain to be solved. They are the development of test content representative of our pluralistic American society, and more appropriate use of tests and test results. It is the contention of many minority educators that standardized achievement and intelligence test content is not reflective of non-white or ethnic group background and heritage.

These minority educators and their supporters in other professions and in civil rights groups have, in some instances, called for a cessation of use of standardized tests with members of racial and

ethnic minority groups. These calls for a moratorium were largely unheeded by the major test developers and users. But, to the dismay of the testing proponents, increasingly vocal criticisms of standardized tests are emanating from the ivory towers of academe, public school classrooms, non-educational professional organizations, and the media.

The last ten years have seen an unprecedented increase in activity among those who question the validity of standardized tests and their attendant demands for accountability on the part of test developers and test users. The names of Hobson, Griggs, Lau, Larry P., DeFunis, and Singleton have joined those of Binet, Otis, Terman, Thorndike, Wechsler, Lennon, and Brigham in test and measurement literature as a consequence of legal action taken by persons who believed themselves to be aggrieved through misuse of test results.

More and more are individuals coming to the conclusion that the entire enterprise of so-called objective standardized ability, aptitude, and achievement tests has been grossly and repeatedly misused. They believe, as well, that the situation warrants the most serious consideration that can be given by law-makers, educators, researchers, test developers and publishers, and test users. Especially needed now, they contend, is a close examination of the problems relating to standardized test use as there is mounting evidence that the influence of the tests on the lives of children and youth, and the educational programs offered to them, is rapidly expanding.

Such scrutiny would prevent the type of situation found in the case of Serna V. Portales Municipal Schools in which Spanish surnamed students alleged they were denied equal educational opportunity in that the school's educational program was tailored for white middle class Anglo children without regard for the educational needs of Spanish speaking children. Expert testimony showed a negative impact on Spanish surnamed students demonstrated by lower IQ scores, achievement scores and language ability. The courts have said repeatedly that race was a dimension which must be considered in interpreting and using the results of performance on standardized tests. They are now extending the same rationale to ethnic heritage as a dimension equally deserving of consideration.

References

Brigham, Carl C. A Study of American Intelligence. Princeton, New Jersey: Princeton University Press, 1923.

Carter, Thomas P. Mexican Americans in School: A History of Educational Neglect. New York: College Entrance Examination Board, 1970.

Kamin, Leon J. The Science and Politics of I.Q. Potomac, Maryland: Lawrence Erlbaum Associates, 1974.

Mercer, Jane R. Labeling the Mentally Retarded: Clinical and Social System Perspectives on Mental Retardation. Berkeley, California: University of California Press, 1973.

Testimony of Dr. Roger T. Lennon as expert witness on psychological testing in the case of Hobson, et al. vs. Hansen, et al. (Washington, D.C. schools) New York: Harcourt, Brace & World, Inc., n.d.

Thorndike, Robert L. Stanford-Binet Intelligence Scale 1972 Norms Tables, 3rd ed., Form 1-B. Boston, Massachusetts: Houghton-Mifflin.